



-52-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANTS: Ji, Hongjun
Rosen, Craig A.
- (ii) TITLE OF INVENTION: Breast Cancer Specific Gene 1
- (iii) NUMBER OF SEQUENCES: 12
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 - (B) STREET: 1100 New York Ave., Suite 600
 - (C) CITY: Washington
 - (D) STATE: DC
 - (E) COUNTRY: USA
 - (F) ZIP: 20005-3934
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 09/017,715
 - (B) FILING DATE: 1998-FEB-03
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/673,284
 - (B) FILING DATE: 28-JUN-96
 - (C) CLASSIFICATION:
- (viii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 60/000,602
 - (B) FILING DATE: 30-JUN-95
 - (C) CLASSIFICATION:
- (ix) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT/US95/08295
 - (B) FILING DATE: 30-JUN-95
 - (C) CLASSIFICATION:
- (x) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 60/037,080
 - (B) FILING DATE: 03-FEB-97
 - (C) CLASSIFICATION:
- (xi) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Steffe, Eric K.
 - (B) REGISTRATION NUMBER: 36,688
 - (C) REFERENCE/DOCKET NUMBER: 1488.0810003
- (xii) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 202-371-2600
 - (B) TELEFAX: 202-371-2540

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 550 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 12..392

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

| | | | | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| CACGAGCCAC | C | ATG | GAT | GTT | TTC | AAG | AAG | GGC | TTC | TCC | ATC | GCC | AAG | AAG | 50 |
| | | Met | Asp | Val | Phe | Lys | Lys | Gly | Phe | Ser | Ile | Ala | Lys | Lys | |
| | | 1 | | | 5 | | | | | 10 | | | | | |
| GGC | GTG | GTG | GGT | GCG | GTG | GAA | AAG | ACC | AAG | CAG | GGG | GTG | ACG | GAA | 98 |
| Gly | Val | Val | Gly | Ala | Val | Glu | Lys | Thr | Lys | Gln | Gly | Val | Thr | Glu | |
| 15 | | | 20 | | | | | | | 25 | | | | | |
| GCT | GAG | AAG | ACC | AAG | GAG | GGG | GTC | ATG | TAT | GTG | GGA | GCC | AAG | ACC | 146 |
| Ala | Glu | Lys | Thr | Lys | Glu | Gly | Val | Met | Tyr | Val | Gly | Ala | Lys | Thr | |
| 30 | | | 35 | | | | | | 40 | | | | 45 | | |
| GAG | AAT | GTT | GTA | CAG | AGC | GTG | ACC | TCA | GTG | GCC | GAG | AAG | ACC | AAG | 194 |
| Glu | Asn | Val | Val | Gln | Ser | Val | Thr | Ser | Val | Ala | Glu | Lys | Thr | Glu | |
| 50 | | | 55 | | | | | | 60 | | | | | | |
| CAG | GCC | AAC | GCC | GTG | AGC | AAG | GCT | GTG | GTG | AGC | AGC | GTC | AAC | ACT | 242 |
| Gln | Ala | Asn | Ala | Val | Ser | Lys | Ala | Val | Val | Ser | Ser | Val | Asn | Thr | |
| 65 | | | 70 | | | | | | 75 | | | | | | |
| GCC | ACC | AAG | ACC | GTG | GAG | GAG | GCG | GAG | AAC | ATC | GCG | GTC | ACC | TCC | 290 |
| Ala | Thr | Lys | Val | Glu | Glu | Ala | Glu | Asn | Ile | Ala | Val | Thr | Ser | Gly | |
| 80 | | | 85 | | | | | | 90 | | | | | | |
| GTG | GTG | CGC | AAG | GAG | GAC | TTG | AGG | CCA | TCT | GCC | CCC | CAA | CAG | GAG | 338 |
| Val | Val | Arg | Lys | Glu | Asp | Leu | Arg | Pro | Ser | Ala | Pro | Gln | Gln | Glu | |
| 95 | | | 100 | | | | | | 105 | | | | | | |
| GAG | GCA | TCC | AAA | GAG | AAA | GAG | GAA | GTG | GCA | GAG | GAG | GCC | CAG | AGT | 386 |
| Glu | Ala | Ser | Lys | Glu | Lys | Glu | Glu | Val | Ala | Glu | Glu | Ala | Gln | Ser | |
| 110 | | | 115 | | | | | | 120 | | | | 125 | | |
| GGA | GAC | TAGAGGGCTA | CAGGCCAGCG | TGGATGACCT | GAAGAGCGCT | CCTCTGCCTT | | | | | | | | | 442 |
| Gly | Asp | | | | | | | | | | | | | | |
| GGACACCATC | CCCTCCTAGC | ACAAGGAGTG | CCCGCCTTGA | GTGACATGCG | GGTGCCCACG | | | | | | | | | | 502 |
| CTCCTGCCCT | CGTCTCCCTG | GACACCCTTG | GCCTGTCCAC | CTGTGCTG | | | | | | | | | | | 550 |

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 127 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Val Phe Lys Lys Gly Phe Ser Ile Ala Lys Lys Gly Val Val
1 5 10 15

Gly Ala Val Glu Lys Thr Lys Gln Gly Val Thr Glu Ala Ala Glu Lys
20 25 30

Thr Lys Glu Gly Val Met Tyr Val Gly Ala Lys Thr Lys Glu Asn Val
35 40 45

Val Gln Ser Val Thr Ser Val Ala Glu Lys Thr Lys Glu Gln Ala Asn
50 55 60

Ala Val Ser Lys Ala Val Val Ser Ser Val Asn Thr Val Ala Thr Lys
65 70 75 80

Thr Val Glu Glu Ala Glu Asn Ile Ala Val Thr Ser Gly Val Val Arg
85 90 95

Lys Glu Asp Leu Arg Pro Ser Ala Pro Gln Gln Glu Gly Glu Ala Ser
100 105 110

Lys Glu Lys Glu Glu Val Ala Glu Glu Ala Gln Ser Gly Gly Asp
115 120 125

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGGGATCCAT GTTTCAAGA AGG

23

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGAAGCTTCT AGTCTCCCCC ACTCTGG

27

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGGGATCCCG ATGTTTCAA GAAGG

25

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGGGTACCCCT AGTCTCCCCC ACTCTGG

27

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGGGATCCGC CACCATGTTT TCAAGAAGG

29

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGGATCCTC AGAAAGCGTA GTCTGGGACG TCGTATGGGT ACTAGTCTCC CCCACTCTGG 60

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGGTACCTC ACTAGTCTCC CCCACTCTGG 30

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3974 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGTACCTAAG TGAGTAGGGC GTCCGATCGA CGGACGCCTT TTTTTGAAT TCGTAATCAT 60

GGTCATAGCT GTTCTCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG 120

CCGGAAGCAT AAAGTGTAAA GCCTGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG 180

CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA 240

TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTCCGCT TCCTCGCTCA 300

CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCAGCGGT ATCAGCTCAC TCAAAGGCC 360

TAATACGGTT ATCCACAGAA TCAGGGGATA ACCGAGGAAA GAACATGTGA GCAAAAGGCC 420

AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTCCAT AGGCTCCGCC 480

CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAC CCGACAGGAC 540

TATAAAAGATA CCAGGCGTTT CCCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC 600

TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA 660

| | |
|---|------|
| GCTCACGCTG TAGGTATCTC AGTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC | 720 |
| ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA | 780 |
| ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG | 840 |
| CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA | 900 |
| GAAGAACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG | 960 |
| GTAGCTCTTG ATCCGGCAA CAAACCACCG CTGGTAGCGG TGGTTTTTT GTTGCAAGC | 1020 |
| AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT | 1080 |
| CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCATGAGA TTATCGTCGA | 1140 |
| CAATTCGCGC GCGAAGGCCGA AGCGGCATGC ATTTACGTTG ACACCATCGA ATGGTGCAAA | 1200 |
| ACCTTTCGCG GTATGGCATG ATAGGCCCG GAAGAGAGTC AATTCAAGGTT GGTGAATGTG | 1260 |
| AAACCAGTAA CGTTATACGA TGTCGCAGAG TATGCCGGTG TCTCTTATCA GACCGTTTCC | 1320 |
| CGCGTGGTGA ACCAGGCCAG CCACGTTCT GCGAAAACGC GGGAAAAAGT GGAAGCGGCG | 1380 |
| ATGGCGGAGC TGAATTACAT TCCCAACCAC GTGGCACAAAC AACTGGCGGG CAAACAGTCG | 1440 |
| TTGCTGATTG GCGTTGCCAC CTCCAGTCTG GCCCTGCACG CGCCGTCGCA AATTGTCGCG | 1500 |
| GCGATTAAT CTCGCGCCGA TCAACTGGGT GCCAGCGTGG TGGTGTGCGAT GGTAGAACGA | 1560 |
| AGCGGCGTCG AAGCCTGTAA AGCGGCGGTG CACAATCTTC TCGCGCAACG CGTCAGTGGG | 1620 |
| CTGATCATTA ACTATCCGCT GGATGACCAAG GATGCCATTG CTGTGGAAGC TGCCTGCACT | 1680 |
| AATGTTCCGG CGTTATTCT TGATGTCTCT GACCAGACAC CCATCAACAG TATTATTTTC | 1740 |
| TCCCATGAAG ACGGTACCGC ACTGGCCGTG GAGCATCTGG TCGCATTGGG TCACCAGCAA | 1800 |
| ATCGCGCTGT TAGCGGGCCC ATTAAGTTCT GTCTCGCGC GTCTCGTCT GGCTGGCTGG | 1860 |
| CATAAAATATC TCACTCGCAA TCAAATTCAAG CCGATAGCGG AACGGGAAGG CGACTGGAGT | 1920 |
| GCCATGTCCG GTTTCAACA AACCATGCAA ATGCTGAATG AGGGCATCGT TCCCAC TGCG | 1980 |
| ATGCTGGTTG CCAACGATCA GATGGCGCTG GCGCAATGC GCGCATTAC CGAGTCCGGG | 2040 |
| CTGCGCGTTG GTGCGGATAT CTCGGTAGTG GGATACGACCG ATACCGAAGA CAGCTCATGT | 2100 |
| TATATCCCGC CGTTAACAC CATCAAACAG GATTTTCGCC TGCTGGGGCA AACCAGCGTG | 2160 |
| GACCGCTTGC TGCAACTCTC TCAGGGCCAG GCGGTGAAGG GCAATCAGCT GTTGCCTGTC | 2220 |
| TCACTGGTGA AAAGAAAAAC CACCCTGGCG CCCAATACGC AAACCGCCTC TCCCCGCGCG | 2280 |
| TTGGCCGATT CATTAATGCA GCTGGCACGA CAGGTTTCCC GACTGGAAAG CGGGCAGTGA | 2340 |
| GCGCAACGCA ATTAATGTAA GTTAGCGCGA ATTGTCGACC AAAGCGGCCA TCGTGCCTCC | 2400 |
| CCACTCCTGC AGTCGGGGG CATGGATGCG CGGATAGCCG CTGCTGGTTT CCTGGATGCC | 2460 |
| GACGGATTG CACTGCCGGT AGAACTCCGC GAGGTCGTCC AGCCTCAGGC AGCAGCTGAA | 2520 |

| | | | | | | |
|--------------|-------------|-------------|-------------|-------------|-------------|------|
| CCAACTCGCG | AGGGGATCGA | GCCC GG GTG | GGCGAAGAAC | TCCAGCATGA | GATCCCCGCG | 2580 |
| CTGGAGGATC | ATCCAGCCGG | CGT CCGGAA | AACGATTCCG | AAGCCCAACC | TTTCATAGAA | 2640 |
| GGCGGC GG TG | GAATCGAAAT | CTCGTGATGG | CAGGTTGGC | GTCGCTTGGT | CGGTCATTTC | 2700 |
| GAACCCCAGA | GTCCC GCTCA | GAAGAACTCG | TCAAGAAGGC | GATAGAAGGC | GATGCGCTGC | 2760 |
| GAATCGGGAG | CGGCGATAACC | GTAAAGCACG | AGGAAGCGGT | CAGCCCATTC | GCCGCCAAGC | 2820 |
| TCTTCAGCAA | TATCACGGGT | AGCCAACGCT | ATGTCCTGAT | AGCGGTCCGC | CACACCCAGC | 2880 |
| CGGCCACAGT | CGATGAATCC | AGAAAAGCGG | CCATTTCCA | CCATGATATT | CGGCAAGCAG | 2940 |
| GCATCGCCAT | GGGT CACGAC | GAGATCCTCG | CCGTCGGGCA | TGCGCGCCTT | GAGCCTGGCG | 3000 |
| AACAGTT CGG | CTGGCGCGAG | CCCCTGATGC | TCTTCGTCCA | GATCATCCTG | ATCGACAAGA | 3060 |
| CCGGCTTCCA | TCCGAGTACG | TGCTCGCTCG | ATGCGATGTT | TCGCTTGGTG | GTCGAATGGG | 3120 |
| CAGGTAGCCG | GATCAAGCGT | ATGCAGCCGC | CGCATTGCAT | CAGCCATGAT | GGATACTTTC | 3180 |
| TCGGCAGGAG | CAAGGTGAGA | TGACAGGAGA | TCCTGCCCG | GCACTTCGCC | CAATAGCAGC | 3240 |
| CAGTCCCTC | CCGCTTCAGT | GACAACGCTG | AGCACAGCTG | CGCAAGGAAC | GCCC GTCGTG | 3300 |
| GCCAGCCACG | ATAGCCGCGC | TGCCTCGTCC | TGCAGTTCAT | TCAGGGCACC | GGACAGGTG | 3360 |
| GTCTTGACAA | AAAGAACCGG | GCGCCCTGC | GCTGACAGCC | GGAACACGGC | GGCATCAGAG | 3420 |
| CAGCCGATTG | TCTGTTGTGC | CCAGTCATAG | CCGAATAGCC | TCTCCACCCA | AGCGGCCGGA | 3480 |
| GAACCTCGGT | GCAATCCATC | TTGTTCAATC | ATGCGAAACG | ATCCTCATCC | TGTCTTTGA | 3540 |
| TCAGATCTTG | ATCCCCTGCG | CCATCAGATC | CTTGGCGGC | AGAAAGCCAT | CCAGTTACT | 3600 |
| TTGCAGGGCT | TCCCCAACCTT | ACCAGAGGGC | GCCCCAGCTG | GCAATTCCGG | TTCGCTTGCT | 3660 |
| GTCCATAAAA | CCGCC CAGTC | TAGCTATCGC | CATGTAAGCC | CACTGCAAGC | TACCTGCTTT | 3720 |
| CTCTTGCGC | TTGCGTTTTC | CCTTGTCCAG | ATAGCCCAGT | AGCTGACATT | CATCCGGG | 3780 |
| CAGCACCGTT | TCTGCGGACT | GGCTTTCTAC | GTGTTCCGCT | TCCTTAGCA | GCCCTTGCGC | 3840 |
| CCTGAGTGCT | TGCGGCAGCG | TGAAGCTTAA | AAA ACTGCAA | AAA ATAGTTT | GA CTTGTGAG | 3900 |
| CGGATAACAA | TTAAGATGTA | CCCAATTGTG | AGCGGATAAC | AATTCACAC | ATTAAAGAGG | 3960 |
| AGAAATTACA | TATG | | | | | 3974 |

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AAGCTTAAAA AACTGCAAAA AATAGTTGA CTTGTGAGCG GATAACAATT AAGATGTACC 60
CAATTGTGAG CGGATAACAA TTTCACACAT TAAAGAGGAG AAATTACATA TG 112

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 755 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCCGCTGCGG CAGACTCGAG CCAGCTCAAG CCCGCAGCTC GCAGGGAGAT CCAGCTCCGT 60
CCTGCCTGCA GCAGCCAACC CTGCACACCC ACCATGGATG TTTCAAGAAC GGCTTCTCCA 120
TCGCCAAGGA GGGCGTGGTG GGTGCGGTGG AAAAGACCAA GCAGGGGTG ACGGAAGCAG 180
CTGAGAACAG CAAGGAGGG GTCATGTATG TGGGAGCCAA GACCAAGGAG AATGTTGTAT 240
GTACAGAGCG TGACCTCAGT GGCGAGAAC ACCAAGGAGC AGGCCAACGC CGTGAGCAAG 300
GCTGTGGTGA GCAGCGTCAA CACTTGGCCA CCAAGACCGT GAGGAGGCGG AGAACATCGC 360
GGTCACTCCG GGTGTGCGCA AGGAGGATTA GGCCATTCCC CCCAACAGGA GGGTGAGGCA 420
TCAAGAACAG AAGGGCAGGC AGAGTGGGG AGACTAGAGG GCTACAGGCC AGCTTGGATG 480
ACCTGAAGAG CGCTCCTCTG CCTTGGGACA CCATCCCCTC CTAGCACAAG GAGTGCCCGC 540
TTTGAGTGGG CATGCGGCTG TCCCACGTTG CTGCCCTCGT TTTCCCTGGG CCACCTTGGC 600
CTGTCCAAGT GTGCTGTTGC AACCAACTTA ATTGCCTTCC TTGGGCCCA ACCAACTTTT 660
TGGTTCTTT TGACCCATTG ATGTTGTTG TGAATTTTT TTTAAAAGA TTTCAAATAA 720
AATTTGGGCC CATTTTTAA AAAAAAAA AAAAA 755

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANTS: Ji, Hongjun
Rosen, Craig A.
- (ii) TITLE OF INVENTION: Breast Cancer Specific Gene 1
- (iii) NUMBER OF SEQUENCES: 11
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 - (B) STREET: 1100 New York Ave., Suite 600
 - (C) CITY: Washington
 - (D) STATE: DC
 - (E) COUNTRY: USA
 - (F) ZIP: 20005-3934
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: TO BE ASSIGNED
 - (B) FILING DATE: HEREWITH
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/673,284
 - (B) FILING DATE: 28-JUN-96
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 60/000,602
 - (B) FILING DATE: 30-JUN-95
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT/US95/08295
 - (B) FILING DATE: 30-JUN-95
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 60/037,080
 - (B) FILING DATE: 03-FEB-97
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Steffe, Eric K.
 - (B) REGISTRATION NUMBER: 36,688
 - (C) REFERENCE/DOCKET NUMBER: 1488.0810003
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 202-371-2600
 - (B) TELEFAX: 202-371-2540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 550 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 12..392

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

| | | | | | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| CACGAGCCAC | C | ATG | GAT | GTT | TTC | AAG | AAG | GGC | TTC | TCC | ATC | GCC | AAG | AAG | | 50 |
| | | Met | Asp | Val | Phe | Lys | Lys | Gly | Phe | Ser | Ile | Ala | Lys | Lys | | |
| | | 1 | | | 5 | | | | | | 10 | | | | | |
| GGC | GTG | GTG | GGT | GCG | GTG | GAA | AAG | ACC | AAG | CAG | GGG | GTG | ACG | GAA | GCA | 98 |
| Gly | Val | Val | Gly | Ala | Val | Glu | Lys | Thr | Lys | Gln | Gly | Val | Thr | Glu | Ala | |
| | | 15 | | | 20 | | | | | | 25 | | | | | |
| GCT | GAG | AAG | ACC | AAG | GAG | GGG | GTC | ATG | TAT | GTG | GGA | GCC | AAG | ACC | AAG | 146 |
| Ala | Glu | Lys | Thr | Lys | Glu | Gly | Val | Met | Tyr | Val | Gly | Ala | Lys | Thr | Lys | |
| | | 30 | | | 35 | | | | 40 | | | | 45 | | | |
| GAG | AAT | GTT | GTA | CAG | AGC | GTG | ACC | TCA | GTG | GCC | GAG | AAG | ACC | AAG | GAG | 194 |
| Glu | Asn | Val | Val | Gln | Ser | Val | Thr | Ser | Val | Ala | Glu | Lys | Thr | Lys | Glu | |
| | | 50 | | | 55 | | | | 60 | | | | | | | |
| CAG | GCC | AAC | GCC | GTG | AGC | AAG | GCT | GTG | GTG | AGC | AGC | GTC | AAC | ACT | GTG | 242 |
| Gln | Ala | Asn | Ala | Val | Ser | Ala | Val | Val | Val | Ser | Ser | Val | Asn | Thr | Val | |
| | | 65 | | | 70 | | | | 75 | | | | | | | |
| GCC | ACC | AAG | ACC | GTG | GAG | GAG | GCG | GAG | AAC | ATC | CCG | GTC | ACC | TCC | GGG | 290 |
| Ala | Thr | Lys | Thr | Val | Glu | Ala | Glu | Ala | Glu | Asn | Ile | Ala | Val | Thr | Ser | |
| | | 80 | | | 85 | | | | | | 90 | | | | | |
| GTG | GTG | CGC | AAG | GAG | GAC | TTG | AGG | CCA | TCT | GCC | CCC | CAA | CAG | GAG | GGT | 338 |
| Val | Val | Arg | Lys | Glu | Asp | Leu | Arg | Pro | Ser | Ala | Pro | Gln | Gln | Glu | Gly | |
| | | 95 | | | 100 | | | | 105 | | | | | | | |
| GAG | GCA | TCC | AAA | GAG | AAA | GAG | GAA | GTG | GCA | GAG | GAG | GCC | CAG | AGT | GGG | 386 |
| Glu | Ala | Ser | Lys | Glu | Lys | Glu | Glu | Val | Ala | Glu | Glu | Ala | Gln | Ser | Gly | |
| | | 110 | | | 115 | | | | 120 | | | | 125 | | | |
| GGA | GAC | TAGAGGGCTA | CAGGCCAGCG | TGGATGACCT | GAAGAGCGCT | CCTCTGCCTT | | | | | | | | | | 442 |
| Gly | Asp | | | | | | | | | | | | | | | |
| GGACACCATC | CCCTCCTAGC | ACAAGGAGTG | CCCGCCTTGA | GTGACATGCG | GGTGCCCACG | | | | | | | | | | | 502 |
| CTCCTGCCCT | CGTCTCCCTG | GACACCCTTG | GCCTGTCCAC | CTGTGCTG | | | | | | | | | | | | 550 |

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 127 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Val Phe Lys Lys Gly Phe Ser Ile Ala Lys Lys Gly Val Val
1 5 10 15

Gly Ala Val Glu Lys Thr Lys Gln Gly Val Thr Glu Ala Ala Glu Lys
20 25 30

Thr Lys Glu Gly Val Met Tyr Val Gly Ala Lys Thr Lys Glu Asn Val
35 40 45

Val Gln Ser Val Thr Ser Val Ala Glu Lys Thr Lys Glu Gln Ala Asn
50 55 60

Ala Val Ser Lys Ala Val Val Ser Ser Val Asn Thr Val Ala Thr Lys
65 70 75 80

Thr Val Glu Glu Ala Glu Asn Ile Ala Val Thr Ser Gly Val Val Arg
85 90 95

Lys Glu Asp Leu Arg Pro Ser Ala Pro Gln Gln Glu Gly Glu Ala Ser
100 105 110

Lys Glu Lys Glu Glu Val Ala Glu Glu Ala Gln Ser Gly Gly Asp
115 120 125

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGGGATCCAT GTTTCAAGA AGG

23

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGAAGCTTCT AGTCTCCCC ACTCTGG

27

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGGGATCCCG ATGTTTCAA GAAGG

25

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGGGTACCCCT AGTCTCCCCC ACTCTGG

27

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGGGATCCGC CACCATGTTT TCAAGAAGG

29

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGGATCCTC AGAAAGCGTA GTCTGGGACG TCGTATGGGT ACTAGTCTCC CCCACTCTGG 60

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGGTACCTC ACTAGTCTCC CCCACTCTGG 30

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3974 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGTACCTAAG TGAGTAGGGC GTCCGATCGA CGGACGCCCTT TTTTTGAAT TCGTAATCAT 60

GGTCATAGCT GTTCTCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG 120

CCGGAAGCAT AAAGTGTAAA GCCTGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG 180

CGTTGCGCTC ACTGCCGCT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA 240

TCGGCCAACG CGCGGGGAGA GGCGGTTGC GTATTGGCG CTCTTCCGCT TCCTCGCTCA 300

CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG 360

TAATACGGTT ATCCACAGAA TCAGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC 420

AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTCCAT AGGCTCCGCC 480

CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAC CCGACAGGAC 540

TATAAAGATA CCAGGCCTTT CCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC 600

TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA 660

GCTCACGCTG TAGGTATCTC AGTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC 720

ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA 780

ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGGCCAC TGGTAACAGG ATTAGCAGAG 840

CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA 900
GAAGAACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG 960
GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGTTTTTTT GTTGCAAGC 1020
AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT 1080
CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATT TTGATGAGA TTATCGTCGA 1140
CAATTGCGC GCGAAGGCAG AGCGGCATGC ATTACGTTG ACACCATCGA ATGGTGCAA 1200
ACCTTCGCG GTATGGCATG ATAGGCCCG GAAGAGAGTC AATTCAAGGT GGTGAATGTG 1260
AAACCAGTAA CGTTATACGA TGTCGCAGAG TATGCCGGTG TCTCTTATCA GACCGTTCC 1320
CGCGTGGTGA ACCAGGCCAG CCACGTTCT GCGAAAACGC GGGAAAAAGT GGAAGCGCG 1380
ATGGCGGAGC TGAATTACAT TCCCAACCGC GTGGCACAAC AACTGGCGGG CAAACAGTCG 1440
TTGCTGATTG GCGTGGCCAC CTCCAGTCTG GCCCTGCACG CGCCGTCGCA AATTGTCGCG 1500
GCGATTAAAT CTCGCGCCGA TCAACTGGGT GCCAGCGTGG TGTTGTCGAT GGTAGAACGA 1560
AGCGCGTCG AAGCCTGTAA AGCGCGGTG CACAATCTTC TCGCGCAACG CGTCAGTGG 1620
CTGATCATTA ACTATCCGCT GGATGACCAG GATGCCATTG CTGTGGAAGC TGCCTGCACT 1680
AATGTTCCGG CGTTATTCT TGATGTCCT TGACAGACAC CCATCAACAG TATTATTTTC 1740
TCCCATGAAG ACGGTACGCG ACTGGCGTG GAGCATCTGG TCGCATTGGG TCACCAGCAA 1800
ATCGCGCTGT TAGCGGGCCC ATTAAGTTCT GTCTCGCGC GTCTGCGTCT GGCTGGCTGG 1860
CATAAATATC TCACTCGCAA TCAAATTCAAG CCGATAGCGG AACGGGAAGG CGACTGGAGT 1920
GCCATGTCCG GTTTCAACA AACCATGCAA ATGCTGAATG AGGGCATCGT TCCCAC TGCG 1980
ATGCTGGTTG CCAACGATCA GATGGCGCTG GGCGCAATGC GCGCCATTAC CGAGTCCGGG 2040
CTGCGCGTTG GTGCGGATAT CTCGGTAGTG GGATACGACG ATACCGAAGA CAGCTCATGT 2100
TATATCCCGC CGTTAACAC CATCAAACAG GATTTTCGCC TGCTGGGCA AACCAGCGTG 2160
GACCGCTTGC TGCAACTCTC TCAGGGCCAG GCGGTGAAGG GCAATCAGCT GTTGCCCGTC 2220
TCACTGGTGA AAAGAAAAAC CACCCCTGGCG CCCAATACGC AAACCGCCTC TCCCCGCGCG 2280
TTGGCCGATT CATTAAATGCA GCTGGCACGA CAGGTTTCCC GACTGGAAAG CGGGCAGTGA 2340
GCGCAACGCA ATTAATGTAA GTTAGCGCGA ATTGTCGACC AAAGCGGCCA TCGTGCCTCC 2400
CCACTCCTGC AGTTGGGGGG CATGGATGCG CGGATAGCCG CTGCTGGTTT CCTGGATGCC 2460
GACGGATTG CACTGCCGGT AGAAACTCCGC GAGGTCGTCC AGCTCAGGC AGCAGCTGAA 2520
CCAACTCGCG AGGGGATCGA GCCCGGGGTG GCGAAGAAC TCCAGCATGA GATCCCCGCG 2580
CTGGAGGATC ATCCAGCCGG CGTCCCGGAA AACGATTCCG AAGCCCAACC TTTCATAGAA 2640
GGCGCGGGTG GAATCGAAAT CTCGTGATGG CAGGTTGGGC GTGCGCTGGT CGGTCAATTTC 2700
GAACCCCAAGA GTCCCGCTCA GAAGAACTCG TCAAGAAGGC GATAGAAGGC GATGCGCTGC 2760

| | |
|---|------|
| GAATCGGGAG CGGCGATAACC GTAAAGCACG AGGAAGCGGT CAGCCCATT C GCCGCCAAGC | 2820 |
| TCTTCAGCAA TATCACGGGT AGCCAACGCT ATGTCCTGAT AGCGGTCCGC CACACCCAGC | 2880 |
| CGGCCACAGT CGATGAATCC AGAAAAGCGG CCATTTCCA CCATGATATT CGGCAAGCAG | 2940 |
| GCATGCCAT GGGTCACGAC GAGATCCTCG CCGTCGGGCA TGCGGCCCTT GAGCCTGGCG | 3000 |
| AACAGTTCGG CTGGCGCGAG CCCCTGATGC TCTCGTCCA GATCATCCTG ATCGACAAGA | 3060 |
| CCGGCTTCCA TCCGAGTACG TGCTCGCTCG ATGCGATGTT TCGCTTGGTG GTCGAATGGG | 3120 |
| CAGGTAGCCG GATCAAGCGT ATGCAGCCGC CGCATTGCAT CAGCCATGAT GGATACTTTC | 3180 |
| TCGGCAGGAG CAAGGTGAGA TGACAGGAGA TCCTGCCCG GCACTTCGCC CAATAGCAGC | 3240 |
| CAGTCCCTTC CCGCTTCAGT GACAACGTCG AGCACAGCTG CGCAAGGAAC GCCCGTCGTG | 3300 |
| GCCAGCCACG ATAGCCGC GC TGCCCTCGCC TGCAGTTCAT TCAGGGCACC GGACAGGTG | 3360 |
| GTCTTGACAA AAAGAACCGG GCGCCCCCTGC GCTGACAGGCC GGAACACGGC GGCATCAGAG | 3420 |
| CAGCCGATTG TCTGTTGTGC CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGGA | 3480 |
| GAACCTGCGT GCAATCCATC TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTTTGA | 3540 |
| TCAGATCTTG ATCCCCGTGC CCATCAGATC CTTGGCGGCA AGAAAGCCAT CCAGTTACT | 3600 |
| TTGCAGGGCT TCCCAACCTT ACCAGAGGGC GCCCCAGCTG GCAATTCCGG TTCGCTTGCT | 3660 |
| GTCCATAAAA CCGCCCAGTC TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTT | 3720 |
| CTCTTGCGC TTGCGTTTTC CCTTGTCAG ATAGCCCAGT AGCTGACATT CATCCGGGGT | 3780 |
| CAGCACCGTT TCTGCGGACT GGCTTTTAC GTGTTCCGCT TCCTTAGCA GCCCTGCGC | 3840 |
| CCTGAGTGCT TGCAGCAGCG TGAAGCTTA AAAACTGCAA AAAATAGTTT GACTTGTGAG | 3900 |
| CGGATAACAA TTAAGATGTA CCCAATTGTG AGCGGATAAC AATTTCACAC ATTAAAGAGG | 3960 |
| AGAAATTACA TATG | 3974 |

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

| | |
|--|-----|
| AAGCTTAAAA AACTGAAAAA AATAGTTGA CTTGTGAGCG GATAACAATT AAGATGTACC | 60 |
| CAATTGTGAG CGGATAACAA TTTCACACAT TAAAGAGGAG AAATTACATA TG | 112 |